

### **REMARKS**

In the final Office Action, the Examiner rejected claims 1-3, 9, and 16 under 35 U.S.C. § 102(b) as anticipated by Colbath et al. ("Spoken Documents: Creating Searchable Archives from Continuous Audio," 2000) (which the Examiner identified as "Kubala et al."); rejected claims 4-8, 10, 11, and 17 under 35 U.S.C. § 103(a) as unpatentable over Colbath et al. in view of Liddy et al. (U.S. Patent No. 5,963,940); and rejected claims 18 and 19 under 35 U.S.C. § 102(b) as anticipated by, or in the alternative, under U.S.C. § 103(a) as unpatentable over Colbath et al. Applicants traverse these rejections.<sup>1</sup> Claims 1-19 remain pending, of which claims 12-15 have been withdrawn from consideration by the Examiner.

At the outset, Applicants respectfully submit that the finality of the Office Action, dated February 5, 2007, is improper. In the final Office Action, the Examiner introduces a new ground of rejection for rejecting several claimed features. For example, with regard to claim 1, in the non-final Office Action dated August 15, 2006, the Examiner pointed to column 5, lines 34-48 and column 8, lines 11-16 of Colbath et al. as allegedly disclosing "determining whether the topics are associated with at least half of the documents in the clusters" (non-final Office Action, pg. 4). In the final Office Action dated February 5, 2007, the Examiner pointed to only column 14, lines 1-26 of Colbath et al. as allegedly disclosing the same feature (final Office Action, pg. 3). With further respect to claim 1, the Examiner pointed to column 5, lines 21-32 of Colbath et al. as allegedly disclosing "forming labels for the clusters from the cluster lists" in the non-final Office Action dated

---

<sup>1</sup> As Applicants' remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicants' silence as to assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art, motivation to combine reference, assertions as to dependent claims, etc.) is not a concession by Applicants that such assertions are accurate or such requirements have been met, and Applicants reserve the right to analyze and dispute such assertions/requirements in the future.

August 15, 2006 (non-final Office Action, pg. 4) and only to column 15, lines 27-37 of Colbath et al. as allegedly disclosing the same feature in the final Office Action dated February 5, 2007 (final Office Action, pg. 3).

M.P.E.P. § 706.07(a) recites "second or any subsequent actions on the merits shall be final, except where the examiner introduces a new ground of rejection that is neither necessitated by applicant's amendment of the claims nor based on information submitted in an information disclosure filed during the period set forth in 37 CFR 1.97(c) with the fee set forth in 37 CFR 1.17(p)" (emphasis added). Since Applicants did not amend claim 1 in the Amendment of November 15, 2006, Applicants' Amendment filed November 15, 2006 could not have necessitated the Examiner's application of a new ground of rejection. Accordingly, Applicants submit that the finality of the Office Action, dated February 5, 2007, is improper. Withdrawal of the finality of the Office Action, dated February 5, 2007, is respectfully requested.

*REJECTION UNDER 35 U.S.C. § 102(b)*

In paragraph 8 of the final Office Action, the Examiner rejected claims 1-3, 9, and 16 under 35 U.S.C. § 102(b) as allegedly anticipated by Colbath et al. Applicants respectfully traverse the rejection.

A proper rejection under 35 U.S.C. § 102 requires that a single reference teach every aspect of the claimed invention. Any feature not directly taught must be inherently present. In other words, the identical invention must be shown in as complete detail as contained in the claim. See M.P.E.P. § 2131. Colbath et al. does not disclose or suggest the combination of features recited in claims 1-3, 9, and 16.

Independent claim 1, for example, is directed to a method of creating labels for clusters of documents. The method comprises identifying topics associated with the documents in the clusters; determining whether the topics are associated with at least half of the documents in the clusters; adding ones of the topics that are associated with at least half of the documents in the clusters to cluster lists; and forming labels for the clusters from the cluster lists.

Colbath et al. does not disclose the combination of features recited in claim 1. For example, Colbath et al. does not disclose clusters of documents. Colbath et al. mentions the word "cluster" in a few places in the context of clustering speakers together (see, e.g., column 5, lines 21-32), but does not disclose clusters of documents.

Because Colbath et al. does not disclose clusters of documents, Colbath et al. cannot disclose determining whether topics are associated with at least half of the documents in the clusters, as further recited in claim 1. The Examiner alleged that Colbath et al. discloses this feature and cited column 14, lines 1-26, of Colbath et al. for support (final Office Action, page 3). Applicants disagree.

At column 14, lines 1-26, Colbath et al. discloses:

One solution to this is to use traditional relevance feedback. The user has the option of specifying an entire story to the query system. When this is done, all the words in the story are fed into the full-text search engine, which returns five documents that use the maximum number of common terms with the seed document.

In the example in Fig. 7, we've given the system the first story in the "Smoking and FDA" query for a relevance feedback operation. The full-text search system has returned five stories. The first one is the seed story, since it has the most terms in common with itself. The second one happens to be the second-ranked story from the boolean query. The remaining three stories, however, are three stories on highly similar topics that weren't found with the boolean query mechanism. It should be emphasized that this mode of search becomes particularly important when the document source is text with errors introduced by a speech recognition system. Because of speech recognition errors, highly relevant documents may fall through the cracks of a boolean search, but are more likely to be found via

relevance feedback since they will contain other words in common that are recognized correctly.

In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or elsewhere, does Colbath et al. disclose determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 3). Applicants cannot understand what the Examiner is alleging. If the Examiner persists with a rejection of the above-identified claim feature based on Colbath et al., Applicants respectfully request that the Examiner provide a reasonable explanation why the Examiner believes that the above-identified sections of Colbath et al. disclose determining whether identified topics are associated with at least half of the documents in the clusters, as required by claim 1.

Colbath et al. also does not disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as further recited in claim 1. Instead, Colbath et al. discloses forming a title for a document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, and cited column 15, lines 27-37, column 8, lines 11-16, column 14, line 29 - column 15, line 25, of Colbath et al. for support (final Office Action, page 3). Applicants disagree.

At column 15, lines 27-37, Colbath et al. discloses:

There is no particular reason that this technique could not be extended to include proper name tagging, marking of new vocabulary words for the recognizer, or identification of new topics for the topic classifier. The latter is particularly important, since it is unlikely that an end-user could find a ready-made set of topics for their own meetings or teleconferences. For some particular problem domains, it may be sufficient to have a small set of topics (3-4 instead of the current 5,500).

In this section, Colbath et al. discloses implementing techniques to include proper name tagging, marking of new vocabulary words for a recognizer, or identification of new topics for a topic classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

At column 8, lines 11-16, Colbath et al. discloses:

The Rough'n'Ready IR system uses a full-text search system developed at BBN which uses an HMM-based model of document retrieval. This system, described in [7], is used in relevance-feedback mode to allow the user of the system to find documents that are similar to an exemplar.

In this section, Colbath et al. discloses an HMM-based model for document retrieval. Nowhere in this section, or elsewhere, does Colbath et al. disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

At column 14, lines 29 - column 15, line 25, Colbath et al. discloses:

#### Annotation

There is no particular reason that the database has to be browsed in a read-only fashion, however. The training data for the Rough'n'Ready indexer is currently fairly static. To annotate more speech data, or additional names for the name spotter, or additional topics for the topic classifier is a separate, offline process using dedicated annotators. However, since the current annotation process is relatively simple and does not require any in-depth linguistic knowledge, it seems logical that the enduser of the archive should be enlisted in helping to provide the training data. This makes sense since it is likely the consumer of the

data will have the most familiar with it, and will be able to provide topics, identify speakers, etc.

The current Rough'n'Ready system includes some basic speaker annotation capabilities. If the user encounters a speaker currently marked as unknown, they can step through a relatively simple wizard that will play segments of data that have been tagged with the same identifier (such as "Male 5") and ask them to confirm that this is the same as the first speaker. Once they have accumulated enough data (three to five minutes), the system trains a new speaker model, and reprocesses the rest of the archive off-line to include the new speaker. It is also possible to add extra training data for speakers that have particularly weak performance, improving their models.

In this section, Colbath et al. discloses that annotators can annotate more speech data, additional names for the name spotter, or additional topics for the topic classifier, and that the current Rough'n'Ready system includes some basic speaker annotation capabilities. Nowhere in this section, or elsewhere, does Colbath et al. disclose forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner also alleged that Colbath et al. discloses a "new identification of new topics for the topic classifier" (final Office Action, page 3). Even assuming, for the sake of argument, that the Examiner's allegation is reasonable (a point that Applicants do not concede), the Examiner has not addressed the feature of forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

In response to the Examiner's arguments on page 8 of the final Office Action that

it should be noted that Kubula discloses the speaker identification and segmentation system creating paragraph-like units between speakers and clustering archived files with a unique name. Thus, the Kubula's teaching of processing a clustered archived files with unique names is valid to read on the broadly claimed limitation of "identifying topics associated with the documents in the clusters,"

Applicants respectfully disagree. Speaker identification and segmentation allows the system to detect changes between speakers, which is important for correct playback of audio sections of an

archive (Colbath et al., column 5, lines 19-23). Speaker identification has nothing to do with “identifying topics associated with the documents in the clusters,” as recited in claim 1.

Even assuming, for the sake of argument, that speaker identification and segmentation could reasonably be equated to identifying topics associated with the documents in the clusters (a point that Applicants do not concede), nowhere does Colbath et al. disclose or suggest that the “unique name” assigned to the archived files are formed from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1.

The Examiner continues to argue that Colbath et al. discloses the recited features of claim 1, but merely restates the previous rejection without explaining how the cited sections of Colbath et al. disclose the recited features of claim 1. If this rejection is maintained, Applicants respectfully request that the Examiner point out how the cited sections of Colbath et al. can reasonably be construed to disclose the features of claim 1.

For at least these reasons, Applicants submit that claim 1 is not anticipated by Colbath et al. Claims 2 and 3 depend from claim 1 and are, therefore, not anticipated by Colbath et al. for at least the reasons given with regard to claim 1.

Independent claims 9 and 16 recite features similar to (but possibly different in scope from) features recited in claim 1. Claims 9 and 16 are, therefore, not anticipated by Colbath et al. for at least reasons similar to reasons given with regard to claim 1.

*REJECTION UNDER 35 U.S.C. § 103(a)*

In paragraph 10 of the final Office Action, the Examiner rejected claims 4-8, 10, 11, and 17 under 35 U.S.C. § 103(a) as allegedly unpatentable over Colbath et al. in view of Liddy et al. Applicants respectfully traverse the rejection.

Claims 4-8 depend from claim 1, claims 10 and 11 depend from claim 9, and claim 17 depends from claim 16. Without acquiescing in the Examiner's rejections with regard to claims 4-8, 10, 11, and 17, Applicants submit that the disclosure of Liddy et al. does not cure the deficiencies in the disclosure of Colbath et al. identified above with regard to claims 1, 9, and 16. For example, Liddy et al. does not disclose or suggest determining whether identified topics are associated with at least half of the documents in the clusters, or forming labels for the clusters from cluster lists that include topics that are associated with at least half of the documents in the clusters, as required by claim 1 (similar features are recited in claims 9 and 16). Instead, Liddy et al. discloses using headlines from newspaper articles or titles from documents in clusters to form labels for the clusters (column 25, lines 34-35).

Claims 4-8, 10, 11, and 17 are, therefore, patentable over Colbath et al. and Liddy et al., whether taken alone or in any reasonable combination, for at least the reasons given with regard to claims 1, 9, and 16. Claims 4-8, 10, 11, and 17 are also patentable over Colbath et al. and Liddy et al. for reasons of their own.

For example, claim 4 recites assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated. Neither Colbath et al. nor Liddy et al. discloses the combination of features of claim 4.

The Examiner admitted that Colbath et al. does not disclose assigning ranks, but alleged that Liddy et al. discloses assigning ranks and cited column 21, lines 28-52, of Liddy et al. for support (final Office Action, page 5). Applicants disagree.

At column 21, lines 28-52, Liddy et al. discloses:



Matcher 55 matches documents by comparing the documents with the query and assigning each document a similarity score for the particular query. Documents with sufficiently high scores are arranged in ranked order in three folders, according to their relative relevance to the substance of a query. There are a number of evidence sources used for determining the similarity of documents to a query request, including:

Complex Nominals (CNs)\*

Proper Nouns (PNs)\*

Subject Field Codes (SFCs)

Single Terms\*

Text Structure

Presence of Negation

Mandatory requirements

\*CNs, PNs, and Single Terms are collectively called "terms."

Documents are arranged for the user based on a two-tier ranking system. The highest-level ranking mechanism is a system of folders. Documents are placed within folders based on various criteria, such as the presence or absence of mandatory terms. The lower-level ranking mechanism sorts documents within each folder based on criteria such as similarity score, document date assignment, etc.

In this section, Liddy et al. discloses ranking search result documents based on a query match.

Nowhere in this section, or elsewhere, does Liddy et al. disclose or suggest assigning ranks to topics, let alone assigning ranks to the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 4.

For at least these additional reasons, Applicants submit that claim 4 is patentable over Colbath et al. and Liddy et al.

Claim 5 recites ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated. Neither Colbath et al. nor Liddy et al. discloses the combination of features of claim 5.

The Examiner alleged that both Colbath et al. and Liddy et al. disclose these features and cited column 7, lines 7-15 and column 8, lines 8-26, of Colbath et al., and column 24, line 56 - column 25, line 2, of Liddy et al. for support (final Office Action, page 6). Applicants disagree.

At column 7, lines 7-15, Colbath et al. discloses:

Topic samples are taken from a sliding 200-word window across the transcribed text. Runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. The current set of approximately 5,500 topics come from an outside vendor, and apply specifically to broadcast news.

In this section, Colbath et al. discloses that runs of similar high-ranking topics are combined to create *story* boundaries that give the user a high-level view of the data being shown, as well as providing a document model for information retrieval. Nowhere in this section does Colbath et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

Column 14, lines 8-26 have been reproduced above. This section of Colbath et al. discloses the results of a full-text search system when the system is given a query for a relevance feedback operation. Nowhere in this section does Colbath et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

At column 24, line 56 - column 25, line 2, Liddy et al. discloses:

The matching of documents to a query organizes documents by matching scores in a ranked list. The total number of presented documents can be selected by the user, the system can determine a number using the Recall Predictor (RP) function, or, in the absence of user input, the system will retrieve all documents with a non-zero score. Note that documents from different sources are interfiled and ranked in a single list.

The RP filtering function is accomplished by means of a multiple regression formula that successfully predicts cut-off criteria on a ranked list of relevant documents for individual

queries based on the similarity of documents to queries as indicated by the vector matching (and optionally the proper noun matching) scores.

In this section, Liddy et al. discloses matching documents to a search query and organizing the documents by matching scores in a ranked list. Nowhere in this section, or elsewhere, does Liddy et al. even mention ranking topics, let alone ranking the ones of the topics based on a number of the documents with which the ones of the topics are associated, as required by claim 5.

For at least these additional reasons, Applicants submit that claim 5 is patentable over Colbath et al. and Liddy et al. Claims 6-8 depend from claim 5 and are, therefore, patentable over Colbath et al. and Liddy et al., whether taken alone or in any reasonable combination, for at least the reasons given with regard to claim 5.

On page 9 of the final Office Action, the Examiner argues that, “[a]lthough Kubula does not explicitly disclose all the claimed limitations, the feature not disclosed by Kubula is disclosed by Liddy. One can not show non-obviousness by attacking references individually where, as here, the rejection is based on a combination of references.” The Examiner, however, fails to explain how the cited sections of Colbath et al. and Liddy et al. disclose the recited features of claims 4-8, 10, 11, and 17. If the 35 U.S.C. § 103(a) rejection is maintained, Applicants respectfully request that the Examiner explain how the cited sections of Colbath et al. and Liddy et al. disclose the recited features of claims 4-8, 10, 11, and 17.

*REJECTION UNDER 35 U.S.C. § 102(b) OR UNDER 35 U.S.C. § 103(a)*

In paragraph 11 of the final Office Action, the Examiner rejected claims 18 and 19 under 35 U.S.C. § 102(b) as allegedly anticipated by, or in the alternative, as unpatentable over Colbath et al. Applicants respectfully traverse the rejection.

Independent claim 18 is directed to a method for creating labels for clusters of documents. The method comprises identifying topics associated with the documents in the clusters; determining whether the topics are associated with at least a predetermined number of the documents in the clusters; and generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters.

Colbath et al. does not disclose the combination of features recited in claim 18. For example, Colbath et al. does not disclose determining whether identified topics are associated with at least a predetermined number of the documents in the clusters. The Examiner alleged that Colbath et al. discloses this feature and cited column 14, lines 1-26 of Colbath et al. for support (final Office Action, page 7). Applicants disagree.

Column 14, lines 1-26, of Colbath et al. has been reproduced above. In this section, Colbath et al. discloses the use of traditional relevance feedback involving feeding all the words in a story into a full-text search engine, which returns five documents that use the maximum number of common terms with the seed document. Nowhere in this section, or elsewhere, does Colbath et al. disclose determining whether identified topics are associated with at least a predetermined number of the documents in the clusters, as required by claim 18.

The Examiner also alleged that "three out of five search results contain highly similar topics for the cluster group" (final Office Action, page 7). Applicants cannot understand what the Examiner is alleging. If the Examiner persists with a rejection of the above-identified claim feature based on Colbath et al., Applicants respectfully request that the Examiner provide a reasonable explanation why the Examiner believes that the above-identified sections of Colbath et al. disclose

determining whether identified topics are associated with at least a predetermined number of the documents in the clusters, as required by claim 18.

Colbath et al. also does not disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as further recited in claim 18. Instead, Colbath et al. discloses forming a title for a document from all of the topics associated with that document (column 10, lines 13-20).

The Examiner alleged that Colbath et al. discloses generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, and cited column 8, lines 11-16, and column 14, line 8 – column 15, line 37, of Colbath et al. for support (final Office Action, page 7). Applicants disagree.

Column 8, lines 11-16, of Colbath et al. is reproduced above. In this section, Colbath et al. discloses an HMM-based model for document retrieval. Nowhere in this section, or elsewhere, does Colbath et al. disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

Column 14, line 8 – column 15, line 37, of Colbath et al. is reproduced above. In this section, Colbath et al. discloses that annotators can annotate more speech data, additional names for the name spotter, or additional topics for the topic classifier. This section of Colbath et al. also discloses including proper name tagging, marking of new vocabulary words for a recognizer, and identification of new topics for a topic classifier. Nowhere in this section, or elsewhere, does Colbath et al. disclose generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

The Examiner also alleged that Colbath et al. discloses a "new identification of new topics for the topic classifier" (final Office Action, page 7). Even assuming, for the sake of argument, that the Examiner's allegation is reasonable (a point that Applicants do not concede), the Examiner has not addressed the feature of generating labels for the clusters using ones of the topics that are associated with the at least a predetermined number of the documents in the clusters, as required by claim 18.

For at least these reasons, Applicants submit that claim 18 is not anticipated by and is patentable over Colbath et al. Claim 19 depends from claim 18 and, therefore, is not anticipated by and is patentable over Colbath et al. for at least the reasons given with regard to claim 18.

#### *CONCLUSION*

In view of the foregoing amendments and remarks, Applicants respectfully request reconsideration and allowance of pending claims 1-11 and 16-19.

If the Examiner believes that the application is not now in condition for allowance, Applicants respectfully request that the Examiner contact the undersigned to discuss any outstanding issues.

Application No. 10/685,479  
Response dated May 7, 2007  
After Final Office Action of February 5, 2007

Docket No.: BBNT-P01-199

Applicants believe no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 06-1075, under Order No. BBNT-P01-199 from which the undersigned is authorized to draw.

Dated: May 7, 2007

Respectfully submitted,

By 

Edward A. Gordon

Registration No.: 54,130

FISH & NEAVE IP GROUP, ROPES & GRAY  
LLP

One International Place  
Boston, Massachusetts 02110  
(617) 951-7000  
(617) 951-7050 (Fax)  
Attorneys/Agents For Applicant